PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C07K 16/00 | A2

(11) International Publication Number: **WO 00/15665**

(43) International Publication Date: 23 March 2000 (23.03.00)

(21) International Application Number: PCT/DK99/00484

(22) International Filing Date: 14 September 1999 (14.09.99)

(30) Priority Data:
PA 1998 01155 14 September 1998 (14.09.98) DK

(71)(72) Applicants and Inventors: PEDERSEN, Lars, Østergaard [DK/DK]; Ved Vigen 14, DK–2400 Copenhagen NV (DK). BUUS, Søren [DK/DK]; Stenmaglevej 29, DK–2700 Brønshøj (DK).

(74) Agent: PLOUGMANN, VINGTOFT & PARTNERS A/S; Sankt Annæ Plads 11, P.O. Box 3007, DK–1021 Copenhagen K (DK).

(81) Designated States: AE, AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), EE, EE (Utility model), ES, FI, FI (Utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: A METHOD OF PRODUCING A FUNCTIONAL IMMUNOGLOBULIN SUPERFAMILY PROTEIN

(57) Abstract

The present invention relates to a process of producing a functional immunoglobulin superfamily protein, which has at least one disulphide bond when functional, the process comprising the steps of providing a bacterial cell comprising a gene coding for the protein, the gene is expressible in said cell, cultivating the cell under conditions where the gene is expressed, isolating the protein from the cell without reducing it, and subjecting the isolated protein to a folding treatment. Preferably, the immunoglobulin superfamily protein is selected from the group consisting of antibodies, immunoglobulin variable (V) regions, immunoglobulin constant (C) regions, immunoglobulin light chains, immunoglobulin heavy chains, CD1, CD2, CD3, Class I and Class II histocompability molecules, $\beta_2$microglobulin (β_2m), lymphocyte function associated antigen–3 (LFA–3) and FcγRIII, CD7, CD8, Thy–1 and Tp44 (CD28), T cell receptor, CD4, polyimmunoglobulin receptor, neuronal cell adhesion molecule (NCAM), myelin associated glycoprotein (MAG), P myelin protein, carcino–embryonic antigen (CEA), platelet derived growth factor receptor (PDGFR), colony stimulating factor–1 receptor, $\alpha\beta$–glycoprotein, ICAM (intercellular adhesion molecule), platelet and interleukins. Important embodiments of the invention is a stable peptide free MHC protein obtainable by a process of the invention and a kit comprising a MHC class I heavy chain and a β_2m allowing the recipient to produce and measure or detect a functional MHC class I protein to which a peptide, which is capable of binding to said MHC class I protein, can be added leading to the generation of a functional MHC class I protein.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

A METHOD OF PRODUCING A FUNCTIONAL IMMUNOGLOBULIN SUPERFAMILY PROTEIN

BRIEF DISCLOSURE OF THE INVENTION

The specificity and reactivity of the human immune system is governed by MHC (in humans HLA) molecules. The function of the HLA is to select and present antigenic peptides to immune T cells. One could say that the immune system views the world through the eyes of the MHC, and that any rational approach to immune manipulations must take MHC into consideration. Such rational approaches should have many scientifical, practical or clinical uses. To exploit these potentials we have devised a method which generates recombinant MHC molecules of high purity in a novel, yet simple, robust and cheap manner. These recombinant molecules are functionally fully active as peptide binders and T cell stimulators. They can be generated in two distinct forms: a) as a fully mature, peptide filled moiety, which is extremely stable and T cell stimulatory, and b) as a partially mature, peptide free moiety ("empty" MHC molecules), which is reasonably stable and readily peptide receptive. It should be noted that the latter result corrects the current misconception that "empty" MHC molecules are non-existent, or at least extremely labile.

All other recombinant or non-recombinant MHC production methods known to us are much more cumbersome and/or generates products of limited efficiency and purity. The complexity and poor stability of MHC molecules makes it very difficult to generate pure and peptide receptive MHC molecules and consequently, it is very difficult to generate predetermined and pure peptide-MHC complexes. As natural molecules from natural sources, the production of MHC molecules is tedious and of low yield. Furthermore, they can only be purified by cumbersome methods leading to preparations where the MHC is preoccupied by a large gamesh of different peptides and sometimes also contaminated by other MHC haplotypes the net result being grossly contaminated preparations.

Compounding the problems in generating MHC molecules is the extremely polymorphism of the MHC locus. In the human population more than 400 different HLA-A, HLA-B and HLA-C alleles exist, and more that 200 HLA-D alleles exist. This diversity has an immunological purpose, but is a practical obstacle to MHC production because many different MHC's need to be generated and individually optimised, validated, characterised, stored etc.

A recombinant expression system might have several advantages. One could potentially obtain a higher yield, an easy purification scheme, and the molecules could be labelled to homogeneity with one particular peptide. A major obstacle of such an approach, however, is that the correctly folded MHC structure is a rather complicated structure consisting of three components (a heavy chain, a light chain (beta2-microglobulin, β_2m) and a peptide). The full stability of MHC is only acquired when all three components are together. In particular the heavy chain is extremely unstable in the absence of the two other components. Thus, it is difficult to produce, handle and store isolated heavy chain without the β_2m and peptide, and it is therefore difficult to generate MHC molecules, which are readily available for binding of any peptide of the experimentors choice. Because of the limited stability, isolated MHC molecules rapidly aggregate and are lost leading to extremely poor final yields.

This patent application describes a general approach to generate recombinant MHC molecules in large amounts and of any allele specificity desired. The method also achieves for the first time truly empty, yet reasonably stable, MHC molecules - and the results show that the binding characteristics of these empty molecules deviate from those obtained with MHC generated in the past. It is a reasonable expectation that these empty molecules are more relevant to physiological MHC binding since they reflect *de novo* binding, whereas previously used MHC molecules have reflected more artificial exchange reactions. In addition to the improved MHC quality, the novel production scheme is easy and robust, readily adaptable to most (probably all) MHC molecules, and it has a high yield of pure and fully functional MHC molecules. These recombinant molecules are extremely potent as the peptide binding activity of as little as 2-4 ng isolated MHC heavy chain can be detected, and the T cell binding capacity of as little as 150 ng peptide/MHC can be detected. They are extremely active with an affinity, which is similar to the one measured for natural occurring molecules, a much faster association rate - and fully available for binding (i.e. without endogenous peptides). We have been able to store these molecules and maintain their activity over many months. The implications are both analytical and therapeutical.

Finally, the methods described have successfully been applied to other (non-MHC) molecules such as components of the CD3 complex. In fact, the methods disclosed would be of use in any protein production scheme (be it recombinant or not, be it in prokaryotes or

in eukaryotes) where the protein at some point during the production is solvated by denaturation, or unfolding (the purpose could be to dissolve protein aggregates, to purify the protein etc) necessitating a later renaturation/refolding step. Thus, the method may have a very large field of application. In particular, we speculate that it might be useful for all members of the immunoglobulin superfamily including antibodies, MHC and T cell receptors. It might be useful for the production of all molecules containing at least one cysteine.

Background of the invention

The immune system can be viewed as one of natures bioinformatics systems. It evaluates any substance that enters into our internal environment, determines its nature and decides whether to take action against it. Proteins and peptides are the most important means of obtaining and conveying such immune information. From this point of view, MHC molecules are at the heart of the immune system. MHC molecules are sampling the entire protein metabolism for peptide information and makes this information available for the central recognition unit of the immune system, the T cell. The current patent application is related to a major undertaking to generate accurate methods to determine and predict the function of the MHC (a complete mapping of all human MHC reactivities). Combining the growing genome databases of primary protein sequences of humans and parasites with the precise knowledge of how this immune molecule handles peptide information will lead to new and powerful strategies for epitope prediction. This in turn will improve the possibilities for directed and efficient immune manipulations. The ability to generate recombinant MHC molecules is enabling this large scale MHC/HLA mapping project - and it does also have practical, clinical and scientifical uses of immediate commercial interest as described below.

Introduction

The purpose of the immune system is to protect our body against microbial organisms (e.g. bacteria, virus, parasites) - and perhaps against cancer. Virtually any threat can be eliminated or neutralised by the immune system. To administrate such powers, the immune system must know what to attack, and what not to attack; ideally, foreign matter should be eliminated, while the body itself should be left unharmed. The true hallmark of the immune system is therefore one of specificity i.e. the ability to discriminate between various targets and in particular to distinguish between self and non-self. The specific im-

mune system consists of a large number of cells, or lymphocytes, with a major subdivision into B and T cells representing humoral (antibody) and cellular responses, respectively. Both cells use receptors, which in their genome are encoded in many bits and pieces allowing an enormous recombinatorial receptor diversity. Each B or T cell carries one, and only one, of these receptors and can recognise a tiny part of the universe. All lymphocytes combined, however, can recognise the universe. The overall specificity of the immune system is generated, regulated and coordinated though processes controlling individual lymphocytes. Deleting, or inactivating, a lymphocyte clone removes the corresponding specificity from the repertoire. Activation and propagation of a lymphocyte clone enhances the corresponding specificity - and allows the immune system to respond quickly and strongly should it be exposed to the same antigen again.

B and T cell specificities

B and T cells use entirely different mechanisms to recognise their targets. B cells recognise soluble antigens, and since they can secrete their receptors as antibodies, they can potentially interact with antigen throughout the fluid phase of the extracellular space. In sharp contrast, the T cell receptor is always membrane bound and it only recognises antigen, which is presented on the membrane of a so called antigen presenting cell (APC). In other words, T cell recognition involves a direct physical interaction between two cells, a T cell and an APC. B and T cells also differ with respect to what they recognise. B cells can recognise organic substances of almost any kind, whereas T cells only recognise proteins (as a biological target, proteins are particularly important since they constitute the structural and functional basis of all life). B cells recognise the three dimensional structure of proteins as illustrated by their ability to distinguish between native and denatured proteins. In contrast, T cells can not distinguish between native and denatured proteins. Early on, this led to the idea that T cells recognise altered proteins. We now know that this is true and that T cells only recognise antigenic peptides presented in association with MHC molecules on the surface of APC's. In general, cytotoxic T cells recognise short peptides (8-11-mers) whose amino and carboxy-termini are deeply embedded within the MHC (i.e. the peptide length is restricted). In contrast, helper T cells tend to recognise longer peptides (13-30-mers or longer) with amino and carboxy terminal ends extending out of the MHC (i.e. the peptide length is unrestricted).

Immune responses and MHC restriction

T cells are of particular importance for the induction of immune responses since they determine the reactivity and specificity of the entire immune system, including B cells. It is therefore appropriate to focus our attention on these cells. T cells can only recognise a given antigen, when it is presented in the context of particular MHC molecule. They are "educated" during ontogeny and further activated during the first priming in processes designed to develop T cells carrying receptors specific for a particular antigen-MHC combinations. These T cells will subsequently only recognise the same exact same antigen-MHC combination. This phenomenon is known as "MHC restriction". Another immune phenomenon, that of "responder status", is also determined by the MHC. Individuals of one MHC haplotype will respond to some antigens, and not to others. Other individuals with other MHC haplotypes will respond differently. These two phenomena are of obvious importance for any rational immune manipulation. As mentioned, we now know that they are both controlled by MHC molecules. These molecules, which have specifically evolved for the purpose of antigen presentation. Our current understanding of antigen presentation can be summarised as follows. First, the foreign substance, the antigen, is taken up by APC's. An intracellular pool of antigenic peptides is generated through proteolytic fragmentation of all the protein antigens available to the cell. This pool of peptides is offered to the MHC molecules of the individual and sampled according to length and sequence; some are bound, while others are ignored (the MHC is said to perform "determinant selection". Subsequently, MHC molecules protect the selected peptides against further degradation, transport them to the surface of the APC and display them for T cell scrutiny.

MHC and polymorphism

Two subtypes of MHC exist, MHC class I and II. These subtypes correspond to two subsets of T lymphocytes: 1) CD8+ cytotoxic T cells, which usually recognise peptides presented by MHC class I molecules, and kill infected or mutated cells T cells, and 2) CD4+ helper T cells, which usually recognise peptides presented by MHC class II molecules, and regulate the responses of other cells of the immune system. MHC class I consists of a 43,000 MW transmembrane glycoprotein (the α chain) non-covalently associated with a 12,000 MW non-glycosylated protein (the β chain, also known as β_2-microglobulin). MHC class II has an overall structure similar to MHC class I although the domain distribution is different. Class II consists of two non-covalently associated transmembrane glycoproteins

of approximately 34,000 and 29,000 MW. The detailed structure of MHC class I and II molecules has been solved at the X-ray crystallography level (Björkman et al., 1987). The most interesting part of the MHC structure is the "upper" part which show a unique peptide binding groove consisting of two alpha helixes forming the walls of the groove and eighth beta-pleated sheaths forming the floor of the groove.

The MHC is extremely polymorphic i.e. many different versions (alleles, allotypes) exist in the population, but each individual has only inherited one or two of these (one from the father and one from the mother). It is also polygenic i.e. several MHC encoding loci exist in the genome allowing for simultaneous expression of several isotypes. Importantly, the majority of the polymorphic residues points towards the peptide binding groove affecting its size, shape and functionality (Matsumura et al., 1992). Peptide-MHC interactions are specific, albeit broad, allowing the binding of many unrelated peptides to each MHC allotype (Buus et al., 1987). The polymorphism dictates the specificity of peptide binding and the biological consequence of this is that each individual in the population educates and shapes a unique T cell repertoire.

The generation of MHC specificity

Structurally, the peptide binding site of the MHC forms a groove, which can be subdivided into various pockets, A through F. The majority of the peptide-MHC binding energy involves main chain atoms of the bound peptide (including the termini for MHC class I); features which are common to all peptides (Matsumura et al., 1992). Only the minority of the binding energy involves peptide side chain atoms, however, these interactions are believed to explain the specificity of the MHC. This mechanism explains how the MHC achieves broad specificity, yet high affinity, peptide binding. Functionally, MHC achieves the broad peptide binding specificity through the recognition of "motifs" (Sette et al., 1987). A motif represents important structural requirements needed for peptide binding such as the presence and proper spacing of particular amino acids in anchor positions. Considerable interest has focused on understanding how MHC specificity and motifs are generated, and on characterising the specificity of various MHC molecules. One of the ultimate goals of this effort is to be able to predict peptide binding. It follows from the MHC polymorphism (it has since been proven both structurally and functionally) that each MHC allotype has it own specificity characteristics. So far, these specificities can only be described experimentally.

The description and prediction of MHC specificity

Two fundamentally different, but complementary, approaches are currently used to determine the peptide binding specificity of MHC. One approach consists of sequencing the peptides already bound to MHC molecules of a given allotype (Buus et al., 1988; Falk et al., 1991), whereas the other approach consists of examining which peptides will bind to the MHC (Buus et al., 1986; Olsen et al., 1994). Both approaches have advantages and disadvantages. The sequencing method deals with naturally processed peptide-MHC complexes, however, it arbitrarily assigns important vs. less important vs. non-important residues and it cannot identify negatively interacting residues. It is thus best suited to identify the most dominant of the positively interacting residues. The latter approach - the direct binding method - is quantitative and it allows comparison of binders vs non-binders. It can identify and quantitate both positively and negatively interacting residues. It is perhaps not surprising that the direct binding method yields better predictability (about 70% of the predicted peptides do indeed bind, than the sequencing method (about 30% success) (Kast et al., 1994). It has been demonstrated that accurate predictions of peptide binding require that the fine specificity of the MHC in question is known in detail (sometimes called extended motifs) (Parker et al., 1994; Rupert et al., 1993; Stryhn et al., 1996). However, obtaining such detailed motifs is very labour and resource intensive. Presently, to determine the fine specificity of every MHC molecule of interest large panels of peptides biased towards particular sequences or motifs are routinely analysed (Parker et al., 1994; Rupert et al., 1993). We have recently developed a peptide library based approach, which yields a correct, unbiased and quantitative description of all functionally important MHC binding residues (Stryhn et al., 1996). It is universal since many different MHC molecules can be addressed with the same set of peptide libraries, and assays to test binding can be developed for any MHC molecule. Conveniently, this approach significantly reduces the experimental set-up and the subsequent data handling and should therefore ease the complete mapping of all MHC class I specificities. This unbiased peptide library based approach does also leads to improved peptide binding predictions (Stryhn et al., 1996). The success of the predicting algorithm implies that MHC class I binding can be viewed largely as the result of an array of independently acting sub-specificities.

The generation of recombinant MHC molecules

Bacteria as production vehicles for recombinant MHC molecules has been demonstrated by others (Parker and Wiley, 1989). However, being packed in inclusion bodies within the bacteria these MHC molecules have been unavailable for peptide binding. Strategies involving complete denaturation and reduction of these inclusion bodies have been used to extract and solubilize the recombinant MHC molecules (Parker et al., 1992; Parker et al., 1992; Parker and Wiley, 1990). This has in turn necessitated the use of an *in vitro* refolding procedures in the presence of reducing/oxidising agents e.g. glutathion (GSH/GSSG). Other components as L-arginine has been added to prevent aggregation and misfolding. However, the *in vitro* folding faces a major problem in the generation of correctly formed di-sulphide bridges. MHC class I heavy chains contains 4 (in some molecules 5) cysteines. There are several possibilities for mis-paired di-sulphide bridges during such re-folding. The general yield of functional MHC class I using this approach has been reported to be low (about 10 - 20%), and of quite slow kinetics (Garboczi et al., 1992).

DETAILED DISCLOSURE OF THE INVENTION

The present invention relates to a process which has been invented in order to solve the problem of having functional immunoglobulin superfamily proteins expressed in aggregates such as inclusion bodies. In the process of the invention the functional protein may consist of several protein subunits which are generated in the same cell or in different cells. In the latter case, the functional proteins – which may very well be two different kinds of proteins, e.g. a heavy chain of a MHC class I protein and a β2microglobulin - may be combined at the time of folding or later.

In one embodiment, the invention relates to a process of producing a functional immunoglobulin superfamily proteins, which have at least one disulphide bond when functional, the process comprising the steps of

(i) providing a bacterial cell comprising a gene coding for the protein, the gene is expressible in said cell,

(ii) cultivating the cell under conditions where the gene is expressed,

(iii) isolating the protein from the cell under conditions, which do not change the disulphide bonds generated by the cell, and, optionally, purifying the protein,

(iv) subjecting the isolated protein to a folding treatment.

When the term "a" or "an" is used in the present specification and claims, this is meant to be one or more, i.e. at least one.

By the term "a functional protein" is meant an immunoglobulin superfamily protein which is capable of performing at least one of the functions attributed to said protein at least to a substantial degree e.g. as assessed by an in vitro assay. By way of example, "a functional MHC class I protein" is defined a protein which comprises a heavy chain, a light chain (b2m) and a peptide. The heavy chain may be truncated in order to make it soluble in aqueous solution. The peptide is a peptide which can be bound to the MHC protein in question. Such peptides may be found by means of e.g. the direct binding method described in Buus et al., 1986, and Olsen et al., 1994.

By "a functional MHC class II protein" is meant a protein which comprises a complex of two heavy chains (an α and a β chain) and a peptide. The heavy chains may be truncated in order to make the complex soluble in aqueous solution. The peptide is a peptide which can be bound to the MHC protein in question. Such peptides may be found by means of e.g. the direct binding method described in Buus et al., 1986, and Olsen et al., 1994.

Particularly preferred embodiments of the invention are processes wherein the MHC protein is a MHC class I protein selected from the group consisting of a heavy chain, a heavy chain combined with a β2m, and a functional mature MHC class I protein; or a MHC class II protein selected from the group consisting of an α/β dimer and an α/β dimer with a peptide. One important aspect of the invention is a process wherein the produced MHC protein is obtained as a peptide free MHC protein.

By "a peptide free MHC class I protein" is meant a protein which comprises a heavy chain associated with a light chain (b2m) but no peptide. A such protein may also be called an "empty" MHC class I protein.

By "a peptide free MHC class II protein" is meant a protein which comprises a heavy chain complex but no peptide. A such protein may also be called an "empty" MHC class II protein.

The present invention is exemplified with reference to MHC class I proteins, but it envisaged that it may be possible in a similar manner to generate all immunoglobulin superfamily proteins (these are by definition disulfide bonded), i.e. a protein selected from the group consisting of antibodies, immunoglobulin variable (V) regions, immunoglobulin constant (C) regions, immunoglobulin light chains, immunoglobulin heavy chains, CD1, CD2, CD3, Class I and Class II histocompatibility molecules, β2microglobulin (β2m), lymphocyte function associated antigen-3 (LFA-3) and FcγRIII, CD7, CD8, Thy-1 and Tp44 (CD28), T cell receptor, CD4, polyimmunoglobulin receptor, neuronal cell adhesion molecule (NCAM), myelin associated glycoprotein (MAG), P myelin protein, carcinoembryonic antigen (CEA), platelet derived growth factor receptor (PDGFR), colony stimulating factor-1 receptor, αβ-glycoprotein, ICAM (intercellular adhesion molecule), platelet and interleukins. The present inventors have already provided data with respect to several MHC class I molecules, $\beta_2$microglobulin, MHC class II molecules, T cell receptor CD3 chains gamma and epsilon.

Cloning of cDNA encoding the various proteins of interest follows standard procedures e.g. as described in Molecular Cloning (Sambrook, Fritsch and Maniatis, Cold Spring Harbor Press, 1989). Briefly, cDNA is synthesised from appropriate cell lines using commercial cDNA synthesis kits (in casu from Pharmacia). For human work the cells would be derived from the panel of HLA expressing EBV transformed human B cell lines from the 12th International Histocompatibility Workshop Cell Lines Panel Database ("HLA: Genetic diversity of HLA. Functional and Medical Implication", Ed. Dominique Charron, EDK press, 1997, or see http://www.icnet.uk/axp/tia/marsh/ihw.html). For HLA-A*0201 an appropriate cell line would be the IHW 9012. The nucleotide sequence corresponding to the desired MHC/HLA molecule can be found at http://www.anthonynolan.com/HIG/index.html, or at http://www.ncbi.nlm.nih.gov. Using this sequence information oligonucleotide primers can be designed to amplify by the polymerase chain reaction the coding region encompassing amino acids 1 - 274 of the mature MHC/HLA molecule from the appropriate cDNA. A relevant forward and backward primer set for the purpose of amplifying HLA-A*0201 and insert it into the NcoI and HindIII restriction sites of the pET28a expression vector (Novagen, see http://www.novagen.com/vectfram.html. The ligated product is transformed into

the bacteria TOP10F' and selected for kanamycin resistance overnight. Several clones are picked and their plasmids prepared by Wizard miniprep (Promega). The plasmids are used as templates in a polymerase chain reaction using the cloning primers and the amplificate is analysed by electrophoresis in agarose and Ethidium bromide stained. Plasmids which lead to amplificates of the appropriate size is sequenced (ABI 310 sequencer) to identify clones, which contain the desired sequence. These clones are secured and used for the subsequent production. A similar strategy can be used to clone any gene of interest.

It is particularly preferred that the protein is a vertebrate protein, e.g. a human, a murine, a rat, a porcine, a bovine, or an avian protein.

In another embodiment, the invention relates to a process of producing a plurality of functional proteins, where at least one of the proteins is the immunoglobulin superfamily, and the plurality of proteins when functional contains at least one intramolecular or intermolecular disulphide bond, the process comprising the steps of

(i) providing a bacterial cell comprising a plurality of genes coding for one protein each, all genes being expressible in said cell,

(ii) cultivating the cell under conditions where the genes are expressed,

(iii) isolating the proteins from the cell under conditions which do not change the disulphide bonds generated by the cell, and optionally, purifying them,

(iv) subjecting the isolated proteins to a folding treatment.

In this embodiment, the protein can be a fusion protein or may be two separate proteins, i.e. co-expressed proteins.

A further embodiment relates to a process of producing a functional immunoglobulin superfamily protein, which has at least one disulphide bond when functional, the process comprising the steps of

(i) providing a cell comprising a gene coding for the protein, the gene is expressible in said cell, the protein being expressed as an aggregate

(ii) cultivating the cell under conditions where the gene is expressed,

(iii) isolating the protein aggregate from the cell under conditions which do not change the disulphide bonds generated by the cell, and optionally, purifying it,

(iv) subjecting the isolated protein to a folding treatment.

A denatured protein may be present in many different conformations – it does not have a distinctive conformation - whereas a folded protein in aqueous solution is present in one or a few distinct conformations. One of the essential features of this invention is that it avoids the conventional solvation of the protein aggregate effected by denaturation under reducing conditions which leads to a completely unfolded protein. Subsequent refolding in order to generate a correctly folded protein is complicated by the requirement for recreation of the correct disulphide bonds. The method of the invention is different in that it has surprisingly been found that the proteins present in the aggregates e.g. the inclusion bodies seem to be present in a functional form having correct disulphide bonds and the task thus is to have them solubilized without breaking the disulphide bonds. The present inventors have found that the denaturing solvation of the protein must be performed under non-reducing conditions without altering of the redox state of the protein. Using denatured proteins with correct disulphide bonds leads to a simplification of the refolding process which may now be as simple as dilution of e.g. urea without adding a redox couple. The folding may, however, be assisted by other proteins such as chaperones; in case of MHC class I it can be assisted by β2m and/or peptide. Further, the folding treatment according to the invention can for certain proteins, e.g. MHC, be performed essentially in the absence of a redox couple such as GSSG/GSH.

The isolation may be performed by disrupting the cell, separating the aggregates such as inclusion bodies (e.g. by centrifugation), optionally washing, extracting the aggregates (e.g. inclusion bodies) in a denaturing agent (e.g. urea or guanidine-hydrochloride, or by other methods known by a person skilled in the art) leading to extraction of soluble protein. This is a schematic outline of the isolation process of (iii) which may be modified or followed e.g. by a purification step as will be evident to the person skilled in the art. This is

a particularly convenient time, for MHC molecules it is actually preferred to add a step of purification since many non-covalent associated molecules including oligo-peptides can be removed. Such purification may be ion-exchange chromatography, size exclusion chromatography, affinity chromatography, hydrophobic interaction, precipitation, filtration, centrifugation and other methods known by the person skilled in the art.

The folding is initiated by diluting the denaturing agent (e.g. urea) to a point which leads to folding of the protein. Preferably, the folding step in the process of the invention (iv) is performed in an aqueous medium which may comprise at least one buffer compound. The protein may then be subjected to a purification procedure as described above.

In the process of the invention a cell which comprises a gene coding for a heterologous or homologous protein, which gene is expressible in said cell, may be any cell. Preferably, the cell is selected from the group consisting of a bacterial cell, a fungal cell, a yeast cell, an animal cell and a plant cell. More preferably, the cell is a bacterial cell selected from the group consisting of a gram positive bacterium and a gram negative bacterium. In a presently preferred embodiment the gram negative bacterium is E. coli including a strain BL21 or a derivative thereof or a strain XA90 or a derivative thereof.

It is contemplated that another useful cell could be a cell which is genetically modified to have a less reducing intracellular environment than a non-modified cell of the same strain, e.g. the cell has been modified to have a reduced or lacking activity of a thioredoxin reductase or an enzyme having a similar effect on the sulfhydryl reducing potential of the cytoplasm, such as a trxB-mutant. Another useful strain may be a strain which is capable of biotinylating the protein, i.e. which is capable of biotinylating a protein which has a biotinylating sequence. The protein may be modified in vivo or in vitro, e.g. phosphorylated, glycosylated, acetylated, amidated or modified in any other appropriate way.

The expressed protein may be located intracellularly, periplasmatically or extracellularly.

The insertion of the gene coding for the functional protein is carried out by any conventional technique for the introduction into a cell of nucleotide sequences, e.g. by transformation, transfection, or transduction. The gene can be inserted into the chromosome of the host cell typically by means of transposome or by a recombination event, or it can be

introduced episomally by means of appropriate vectors. Suitable vectors for such purposes includes the pET vectors, such as the T7 promoters.

It will be appreciated that the gene can be introduced into the host cell alone or in combination with further nucleotide sequences including sequences that regulate the expression of the gene such as promoter sequences, sequences regulating the promoter, enhancer sequences, sequences coding for repressor substances including antisense RNA, or termination sequences. In order to enhance the amount of protein produced multiple copies of the gene coding for the functional protein can be introduced into the host cell. It is also possible to introduce sequences coding for chaperone proteins or sequences regulating the expression or functionality of native chaperone proteins, or sequences coding for gene products that result in glycosylation of the functional protein. The promoter may be constitutive or inducible.

The process according to the invention will be advantageous in one or more aspect of protein product. It may be that the yield of functional protein produced according to the process relatively to the yield of functional protein obtained under essentially similar conditions but where step (iii) is performed under conditions which do change the disulphide bonds generated by the cell, is increased by at least 10%, such as at least 20%, at least 40%, at least 50%, at least 70%, or at least 100%. Alternatively, it may be that the speed of the process of the invention compared to when step (iii) is performed under conditions, which do change the disulphide bonds generated by the cell, is at least 10% faster, such as at least 20%, at least 40%, at least 50%, at least 70%, or at least 100%. It is contemplated that the increase of speed may in fact be as much as 2 fold, 5 fold, 10 fold, 100 fold or 10000 fold increased. In a presently preferred embodiment, the speed is increased by at least 50 fold. Finally, it may be that the purity of the functional protein produced according to the process relatively to the purity of the resulting functional protein obtained under essentially similar conditions but where step (iii) is performed under conditions, which do change the disulphide bonds generated by the cell, is increased by at least 10%, such as at least 20%, at least 40%, at least 50%, at least 70%, or at least 100%.

With reference to the examples, in particular the MHC protein A2, the folding efficiency may be at least 40%, whereas the MHC protein Db has a folding efficiency which is considered to be even higher, i.e. at least 50%. The % is measured on the active protein immediately after the folding compared to the amount of input protein in question in the

folding process. It is contemplated that when the process according to the invention has been optimised with regard to the protein in question, then at least 25% of the produced immunoglobulin superfamily protein is obtained in a functional form.

Preferably, the protein comprises no unpaired Cysteine residues. However, within the scope of the invention is an embodiment wherein the protein comprises 1 unpaired Cysteine residue. The protein may comprise at least 2, such as at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, or at least 20 Cysteine residues. Preferably, the protein comprises an even number of Cysteine residues. Most likely, the protein is having at most 20, such as at most 14, at most 10, at most 8, at most 5, at most 4, at most 3, or at most 2 Cysteine residues. The protein is preferably capable of having at least 1, such as at least 2, at least 3, at least 4, at least 5, or at least 6 disulphide bonds. Most likely, the protein is capable of having at most 20, such as at most 15, at most 10, at most 8, at most 5, at most 4, at most 3, or at most 2 disulphide bonds.

In a preferred embodiment the gene is a derivative of a naturally occurring gene. The derivative may be obtained by substituting at least one codon which is used more frequently by the host cell than the one originally present where the codon codes for the same amino acid. Generally, the gene is under control of a regulatory DNA sequence not naturally associated with the gene. However, it may also under control of its own promoter. In a presently preferred embodiment the bacterial cell is transformed with an expression vector selected from the group consisting of pET vectors, e.g. T7 promoters. Other vectors will be evident to the person skilled in the art.

A most important aspect of the invention is a stable peptide free MHC protein which is obtainable by a process according to the invention. A stable peptide free MHC protein has not previously been generated by any of the methods within the art. Although it has been claimed (Matsumura et al., 1992) that empty molecules can be obtained from TAP deficient eucaryotic cells such as T2 or insect cells, peptides have been extracted and characterised from such preparations (Wei & Cresswell, 1992; Henderson et al., 1992), i.e. these preparations are not truly empty). One of the uses of the stable peptide free MHC protein is it provides for a highly efficient production of pure homogenous peptide-MHC complexes. By the term "stable" is meant that the heavy chain in isolated form in urea can

be stored for at least 3 months at –20°C at 50% glycerol. The half-life of the complex is actually more than six months. The stability of the functional MHC complex in aqueous solution, i.e. the heavy chain and β2m in 1:1 is being investigated. However, it is known that the half-life of the heavy chain in the presence of an excess of β2m is stable in aqueous solution is about 2 days at 4°C.

Another important aspect of the invention is a kit comprising a MHC class I heavy chain and a β2m allowing the recipient to produce a functional MHC class I protein to which a peptide, which is capable of binding to said MHC class I protein, can be added leading to the generation of a functional MHC class I protein. The MHC proteins will preferably be produced by the method according to the invention. In one embodiment, the kit comprises reagents that will allow the end-user to determine the binding of any peptide of his/her choice using detection systems such as enzyme linked immuno sorbent assay (ELISA), radio immuno assay (RIA), or others know to the person skilled in the art. In another embodiment, the kit comprises an oligomerization of MHC proteins, such as two, three, four or more. In a specific embodiment, the kit comprises a further reagent added as a marker making the kit suitable for diagnostic purposes. The marker is preferably selected from the group consisting of fluorochromes, enzymes, chemiluminescense, and radioactive markers.

A preferred use of the process of the invention is in the manufacturing of MHC, in particular peptide free MHC molecules. A preferred use of a stable peptide free MHC protein is in analysis of the effect of changing an amino acid in the MHC on the binding specificity of said MHC as assessed by an analysis using a peptide library approach be it a synthetic or recombinant library. Such a combination of MHC point mutations followed by specificity analysis by peptide libraries constitutes a novel approach to the examination of structure-function relationships. Further uses of the invention are described in the following.

The use of recombinant MHC I molecules

a) Analysis/diagnostics (in particular by ELISA and/or FACS)

a.1) Quantitation of peptide-MHC interaction. It is of considerable interest to measure the interaction between peptide and MHC. Any putative T cell epitope should be checked for MHC binding, preferably in a quantitative assay. Current methodology for measuring such

interactions is hampered by poorly controlled assay systems, the lack of empty MHC molecules and/or the requirement for labelling at least one of the components. Empty molecules are highly active and easily adaptable to highly controlled biochemical assay systems of the RIA type (Buus et al, 1995). Empty molecules are also adaptable to detection by an ELISA approach. In a preferred embodiment, this ELISA involves pan-specific anti-MHC antibody capture and pan-specific anti-MHC antibody detection allowing highly sensitive, quantitative, non-radioactive detection. Other designs of the ELISA (e.g involving affinity tags) are known to the person skilled in the art.

a.2) Enumeration of specific T cells for quantitative and qualitative characterisation of T cell populations. Mark Davis and coworkers in 1996 (Altman et al., 1996) demonstrated that recombinant, pure peptide-MHC class I complexes could be generated with an added biotinylation signal attached to the C-terminal of the heavy chain. After an enzymatic biotinylation process these complexes could be tetramerized with Phycoreythin labelled Streptavidin. These multimerized, labelled peptide-MHC class I complexes could subsequently be used to label T cells in a peptide-specific, MHC I-restricted manner. The stability of the T cell receptor for peptide-MHC I complexes is generally thought to be too low to effect a stable biochemical binding. However, after tetramerization, the avidity of the multimerized complexes is sufficient to effect bichemical binding. Thus, such tetramers can be used to count by fluorescence activated cell sorter (FACS) analysis of the number of T cells in any given cell suspension. It has subsequently been shown that older methods for counting specific T cells, limiting dilution analysis, is grossly incorrect and underestimates the number of specific T cells. Thus, it has become imperative both from a scientific point of view, but also from a publication point of view, to do "tetramer" analysis. The Davis method has two major problems: it is difficult to produce large amount of pure MHC I molecules, and the biotinylation process is expensive, cumbersome, and in particular, it requires extended incubation at 37°C. Many peptides do remain associated to the MHC I for the time of this latter incubation, and this may explain the variable result even within the same laboratory. The present patent application discloses a method whereby peptide-MHC I complexes can be generated in a fast and efficient process leading to a minimum post-complex formation clean-up. It is envisioned that the MHC I molecules produced according to the present method will be amenable to transport and storage in a way that allows the generation of a commercial kit, which would further allow the final peptide-MHC I complex production to be done in the any non-expert laboratory using any relevant peptide of the end-users choice. In contrast, previous ways of producing

MHC I would require considerable protein and molecular biology knowledge and experience. Finally, many better ways of tagging the MHC I than the enzymatic biotinylation process can be envisioned (Gallimore et al., 1998; Walter et al., 1998).

a.3) Enumeration of specific T cells enabling immune manipulations to be accurately monitored. Any immune manipulation (e.g. vaccination) is in need of an accurate and specific monitoring. The above "tetramer" technology is at this time the golden standard of scientific, clinical and commercial evaluation of immune manipulation.

b) Scientifically

b.1) Functional and structural determination of the specificity of MHC I molecules. MHC I molecules are central players in the generation of all T cell mediated responses. Considerable efforts are aimed at understanding the function and specificity of MHC I molecules. For all these purposes one needs access to functionally active MHC I molecules. Producing MHC I molecules from natural sources have several serious draw-backs (cumbersome, expensive production, yet it yields small amounts of impure MHC I). The method disclosed here allows for an easy, fast and highly efficient production of peptide-MHC I complexes. Previous methods, which were used to produce peptide-MHC I complexes included a step where an excess of peptide was offered to the MHC I under refolding (Garboczi et al., 1992). Thus, the resulting complexes were preoccupied with peptide and therefore not readily available for *de novo* peptide binding. According to the method disclosed here, the MHC I molecules produced can immediately be used for binding of peptide, and will therefore be useful for any analysis including, specificity analysis, of peptide binding. Combining the recombinant MHC I molecules with the recently published peptide library approach (Stryhn et al., 1996), one could examine in detail the specificity of any MHC I molecule including mutated MHC I molecules. It should be stressed that such detailed analysis also leads to improved peptide binding predictions (Stryhn et al., 1996).

b.2) Functional and structural determination of the specificity of T cells. Peptide-MHC I complexes can be generated fast, pure and efficiently by the disclosed method. Such complexes can be used to analyse the structure of the T cell receptor in interaction with said peptide-MHC I complex, and using peptide variants and or MHC I variant it will also be possible to perform a functional analysis of the T cell receptor specificity.